

# Significância clínica: uma revisão do tema

BÁRBARA ANTUNES  
ALEXANDRA OLIVEIRA  
FRANCISCO PIMENTEL  
PEDRO LOPES FERREIRA

Estando a qualidade de vida (QdV) a tornar-se num indicador cada vez mais relevante na forma como o doente e o médico lidam com a doença oncológica, é necessário perceber e avaliar criteriosamente as mudanças de QdV ao longo do tratamento. Tendo um indicador de medida rigoroso e estável, o médico pode avaliar se a mudança de QdV é ou não significativa e se, porventura, se justifica uma alteração da terapêutica e, portanto, da decisão clínica. A interpretação dos resultados dos estudos de QdV tem vindo a ser objecto de escrutínio por parte dos investigadores, pois resultados estatisticamente significativos (que sabemos não serem produto do acaso) não podem obrigatoriamente ser considerados clinicamente significativos. É necessário,

então, uma medida para que se possa identificar a diferença mínima significativa que doente e clínico necessitam conhecer, para proceder a alterações na terapêutica e/ou nas actividades do dia-a-dia. O objectivo deste artigo é proporcionar uma revisão da literatura acerca do conceito de significância clínica na área de oncologia. Espera-se que este artigo possa contribuir para promover, em Portugal, uma melhor compreensão e investigação acerca deste tema, bem como a sua utilização na prática clínica diária, de forma a promover a melhor QdV do doente oncológico.

**Palavras-chave:** oncologia; qualidade de vida; significância clínica.

## 1. Introdução

Apenas com uma avaliação precisa da qualidade de vida (QdV) dos doentes se pode obter informação clínica importante, especialmente na área da oncologia. As alterações na QdV são, de facto, indicadores importantes do impacto de uma nova terapia citotóxica, podem afectar a vontade do doente em continuar o tratamento e podem ajudar a definir uma resposta, na ausência de objectivos ou resultados finais quantificáveis, como uma regressão tumoral. Assim, pode dizer-se que a QdV se está a tornar num indicador cada vez mais relevante da forma como o doente e o médico lidam com a doença maligna. Trata-se de um conceito multidimensional e, por isso, a sua avaliação requer uma abordagem de quantifica-



Barbara Antunes é psicóloga clínica, mestre em Psicologia da Dor. Serviço de Oncologia Médica do Hospital São Sebastião, Santa Maria da Feira.

Alexandra Oliveira é psicóloga clínica, bolseira de doutoramento pela FCT em Ciências da Saúde. Serviço de Oncologia Médica do Hospital São Sebastião, Santa Maria da Feira.

Francisco Pimentel é professor associado convidado. Secção Autónoma de Ciências da Saúde da Universidade de Aveiro.

Pedro Lopes Ferreira é professor associado. Faculdade de Economia da Universidade de Coimbra.

Entregue em Dezembro de 2008

ção de variáveis subjectivas, nomeadamente o bem-estar físico, funcional, social, emocional e espiritual. A complexidade da experiência diária do doente, as alterações frequentes que ocorrem naquelas dimensões e todos os acontecimentos marcantes (vistos pelo doente como tal) desempenham um papel importante na forma como este vive a sua doença. Os métodos estatísticos mais comuns para medir a resposta de um questionário à mudança (mudança relativa, efeito do tamanho, resposta estatística, erro padrão de medida) não dão informação acerca da significância clínica, antes, são descrições quantitativas de mudança. Este trabalho tem como objectivo fazer uma revisão bibliográfica do tema significância clínica na área de oncologia. Espera-se poder contribuir para a promoção, em Portugal, de uma melhor compreensão e investigação acerca deste tema, bem como a sua utilização na prática clínica diária, de forma a conseguir-se uma melhor QdV do doente oncológico.

## 2. Significância clínica

Mais importante do que tratar os resultados da avaliação da QdV pelos métodos estatísticos tradicionais, quantificando-os matematicamente, torna-se imperativo que comecem a ser tratados também na perspectiva do significado clínico que cada acontecimento tem para cada doente em particular. Como em qualquer outro estudo, é essencial que os objectivos da QdV sejam analisados de uma forma estatisticamente rigorosa, até porque a maior parte dos instrumentos de medição utilizados envolve pontuações de múltiplos itens, criando diversas comparações estatísticas potenciais (Movsas, 2004).

Tem havido alguma discordância em relação ao termo correcto a utilizar, no que diz respeito ao conceito que é o objectivo deste trabalho. Por um lado, tem-se utilizado o termo «significância clínica» em que a palavra «clínica», segundo Sloan *et al.* (2002), corresponde a implicações para o doente na tomada de decisão ou na avaliação (êxito ou fracasso) de determinado tratamento. Pode também referir-se ao contexto de pura pesquisa em decidir entre que tratamentos de diferentes ensaios clínicos se demonstra qual o mais bem sucedido. Por outro lado, também se utiliza o conceito de «mudança clinicamente significativa» que nos parece poder gerar confusão com o termo «diferença mínima significativa», uma medida dos métodos de âncoras, tratado mais à frente. Assim, neste trabalho optamos pela expressão «significância clínica».

Quanto à distinção entre «significância clínica» e «significância estatística», esta constitui, em lingua-

gem comum, a distinção entre o que é importante ou grande (tamanho do resultado em si) e a probabilidade de um resultado ocorrer por acaso (Sloan e Dueck, 2004). A significância estatística é determinada pelo valor  $p$  e indica que a alteração na QdV provavelmente não foi causada pelo acaso; a significância clínica refere-se à alteração na QdV suficiente para se detectar uma mudança no curso do tratamento do doente. Para ilustrar a diferença entre estes dois conceitos, considere-se um estudo hipotético que investiga a perda de peso em doentes oncológicos com um determinado tratamento. A perda de peso média de 0,57 kg ( $n = 1000$ , erro padrão = 0,33) é considerada estatisticamente significativa com  $p < 0,0001$ , mas qualquer médico argumentaria que esta diminuição de peso não merece uma alteração no tratamento ou no cuidado ao doente.

Embora uma descoberta estatisticamente significativa sugira que os dados observados não podem ser explicados por flutuações de acaso, há muito tempo que existe consenso de que os valores  $p$  não têm relação directa com a significância clínica (Movsas, 2004). Podem assim levantar-se algumas questões interessantes, como, por exemplo, «se uma alteração de 10 pontos numa escala de 0 a 100 para um item é estatisticamente significativa, isso representará que é também clinicamente significativa?» ou «será que o doente e/ou o médico iriam de facto notar tal alteração de pontuação?».

Pode dizer-se que é uma questão de interpretabilidade: o que muda na pontuação corresponde a um benefício trivial, pequeno, moderado ou grande (Guyatt *et al.*, 2002). Por exemplo, se uma pessoa melhora em cinco pontos a funcionalidade emocional, quer dizer que está mais feliz na companhia da família, terá menos saudades do emprego e não terá de fazer mais medicação anti-depressiva? Ou nenhuma destas mudanças irá ocorrer? Se uma doente crónica do pulmão melhora em cinco pontos a funcionalidade física, será agora capaz de subir um lance de escadas sem limitação? Será capaz de acompanhar o cônjuge numa caminhada e voltar a brincar com os seus netos? Ou permanecerá incapacitada por dispneia? Situações em que resultados estatisticamente significativos são discutíveis como clinicamente significativos ocorrem frequentemente na literatura.

## 3. Diferença mínima significativa

Diferença mínima significativa (DMS) pode ser definida como a diferença mais pequena na pontuação de um domínio de interesse que os doentes percepcionam como importante, quer benéfica quer maléfica,

que levaria o médico a considerar uma alteração nos cuidados ao doente (Jaeschke, Singer e Guyatt, 1989).

Os mesmos autores propõem dois passos no processo de tornar os resultados significativos. O primeiro é compreender o que as mudanças na pontuação significam para o indivíduo e, o segundo, é transformar os resultados de estudos clínicos compreensíveis para os decisores clínicos. A apresentação de alterações da média da QdV (por exemplo, o grupo em que o tratamento melhorou em cinco pontos comparativamente ao grupo de controlo) pode ser errónea. Além disto, a DMS varia por doente e, possivelmente, por grupos de doentes. A proporção de doentes que atingem um determinado grau de benefício e o correspondente número necessário para tratar (NNT) para assegurar que uma pessoa obtenha esse benefício, proporciona um modo mais informativo de apresentar resultados.

Existem, também aqui, duas estratégias distintas para interpretar os resultados das medidas de QdV: uma baseada numa âncora e outra baseada na distribuição. De seguida, cada uma destas categorias é desenvolvida com um pouco mais de pormenor.

### 3.1. Métodos baseados em âncoras

Uma âncora é uma medida independente, ou padrão, que pode, por si mesma, ser interpretada e correlacionada com o instrumento de medição, permitindo à DMS distinguir diferenças triviais de diferenças pequenas mas importantes. Existem múltiplas âncoras no SF-36, como o facto de que apenas 32% dos doentes que pontuaram 40 na escala de função física poderem andar apenas um quarteirão sem dificuldades *versus* 50% daqueles que pontuaram 50 na escala.

Há também diversos factores que fazem com que o conceito de DMS seja útil. Em primeiro lugar, liga a magnitude de alterações às decisões de tratamento na prática clínica. Em segundo, a mais pequena diferença importante que se pretende detectar contribui para o desenho do estudo e para a escolha do tamanho da amostra. Em terceiro, põe ênfase na primazia da perspectiva do doente e, implicitamente, liga essa perspectiva à do médico, permitindo a tomada de decisão conjunta. Por fim, em quarto, o conceito parece ser facilmente entendido por clínicos e investigadores, embora haja ainda pouca experiência com doentes (Spilker, 1996).

A DMS pode também variar entre doentes, como atrás se afirmou. Usando o EORTC QLQ-C30, Osoba *et al.* (1998) pediram aos doentes para preencher este instrumento em vários pontos temporais

diferentes, assim como para graduar a sua percepção de alteração desde o anterior preenchimento do questionário. Avaliaram o funcionamento físico, emocional e social, assim como as escalas de QdV global, cada com uma delas a variar de 0 a 100, e descobriram que, se os resultados variavam entre cinco e dez pontos, os doentes consideravam a sua condição «um pouco melhor ou pior», se variava entre 10 e 20 pontos consideravam-na «moderadamente melhor ou pior» e viam-na como «muito melhor ou pior» se a alteração fosse superior a 20 pontos. A desvantagem deste método é que se baseia na memória dos doentes, solicitando-lhes que se refiram à forma como se sentiam na última vez que tinham preenchido o instrumento. Num outro estudo, King (1996) descobriu que a diferença de cinco ou menos pontos no EORTC QLQ-C30 era considerada uma diferença pequena. No entanto, a definição de uma diferença grande pareceu variar para as diferentes escalas (16 para QoL global, 7 para o funcionamento emocional, 27 para o funcionamento físico). Subsequentemente, Osoba *et al.* (1998) usaram esta abordagem (DMS entre cinco e dez) na interpretação de vários estudos e, num estudo de doentes com astrocitoma anaplásico a fazer quimioterapia, descobriram que os doentes cujas pontuações do EORTC QLQ-C30 melhoraram em mais de dez em três ou mais domínios tinham maior probabilidade (82%) de ter uma resposta total ou parcial do tumor do que uma estabilização ou progressão da doença.

Uma vez estabelecida a DMS, o passo seguinte envolve uma estimativa da proporção de doentes que atingiram este valor, para tornar os resultados mais significativos do ponto de vista clínico, em termos do número de doentes necessários para tratar. Por exemplo, se a DMS está estabelecida em 0,3 e a diferença média de um estudo é de 0,2, não se deveria assumir que esta diferença entre o grupo de controlo e o grupo de tratamento fosse trivial. Antes, a diferença média de 0,2 podia ocorrer se 20% dos doentes tivessem um benefício de 1,0 e 80% não tivessem qualquer benefício, o que corresponderia a um NNT de 5. Claro que tem que se considerar não apenas aqueles que atingiram uma melhoria, mas também a diferença entre grupos no que diz respeito à deterioração. Conforme está representado na *Tabela I*, de entre as abordagens baseadas numa âncora podem distinguir-se as que têm um enfoque na população das que têm um enfoque no indivíduo (Guyatt *et al.*, 2002).

As abordagens focadas numa âncora são análogas à validação de critério, dependem de uma única âncora e estabelecem uma DMS com base na mudança do resultado total, necessitando de dois passos. O primeiro, estabelece a diferença mais pequena que os doentes consideram, em média, ser importante, a

DMS. O segundo passo estima a proporção de doentes que atingiram esse valor. As âncoras para a abordagem com enfoque no indivíduo incluem valores globais de alteração intra-doentes (longitudinal) e entre-doentes (transversal).

As abordagens com enfoque na população são análogas à validação de constructo e dependem de várias âncoras que enquadram a resposta do indivíduo em termos da população total na distribuição (por exemplo, um grupo de doentes com uma pontuação de 40 ter uma mortalidade de 20%). As âncoras para as abordagens baseadas na população incluem o estado funcional de um único item, diagnóstico, sintoma ou gravidade da doença, como, por exemplo, a gravidade da doença poder sustentar âncoras para a interpretação do EORTC QLQ-C30 ou o caso em que os doentes em pior estado de saúde mostraram piores pontuações em variáveis como a presença ou ausência de doença metastática ou prognóstico (Osoba *et al.*, 1998). Os profissionais que tratam estes doentes regularmente poderão considerar estes resultados úteis, na medida em que melhoram a interpretabilidade do instrumento de medição, associam as suas pontuações à perda de emprego (as doenças crónicas afectam bastante a funcionalidade), à utilização de cuidados de saúde (aqueles com uma avaliação da sua saúde mais baixa, farão mais visitas ao médico, utilizarão mais medicação e necessitarão de mais hospitalizações) e à mortalidade (a mortalidade está associada à percepção de deterioração da saúde por parte do doente).

A abordagem de várias âncoras reconhece que as alterações de pontuação que representam efeitos

pequenos, médios ou grandes podem mudar, dependendo da doença e da sua gravidade, ou mesmo do tipo de escala utilizada. Este facto é ao mesmo tempo a sua força, pois evita simplificação enganosa, e a sua fraqueza, pois pode impor um fardo cognitivo excessivo nos doentes e médicos.

### 3.2. Métodos baseados na distribuição

Os métodos baseados na distribuição são distintos dos anteriores no que diz respeito à interpretação dos efeitos, pois têm em conta a distribuição dos resultados (Movsas, 2004). A principal vantagem é que, ao contrário dos métodos baseados numa âncora, pouco esforço é necessário para gerar estes valores, já que há sempre uma medida de variabilidade que pode ser usada. Como limitação fundamental constata-se que as estimativas da variabilidade diferem de estudo para estudo. Por exemplo, se se opta pelo desvio padrão entre-doentes, é necessário ter em conta que depende da heterogeneidade da população a ser estudada: se é extremamente heterogénea, um efeito importante pode ser pequeno em termos de desvio padrão entre-pessoas e, por isso, ser considerado trivial. O mesmo efeito, numa população mais homogénea pode ser grande em termos de desvio padrão entre-pessoas e, assim, ser considerado extremamente importante. O verdadeiro impacto da alteração permanece o mesmo, mas a interpretação muda radicalmente (Movsas, 2004).

Existem, pelo menos, duas maneiras de lidar com este problema. A primeira é seleccionar a variabili-

**Tabela I**  
**Abordagens baseadas em âncoras para estabelecer a interpretabilidade**

	Enfoque no indivíduo (1 âncora)	Enfoque na população (várias âncoras)
Número de âncoras	Uma (semelhante à validação de critério)	Muitas (semelhante à validação do constructo)
Especificação do limiar	Especifica o limiar entre uma mudança importante e uma trivial (DMS)	Em vez do limiar, oferece relações entre a medida a estudar e várias âncoras
Passos na aplicação	Dois passos: estabelecer a DMS e examinar proporção que atingiu este valor	Um passo: apresentar as diferenças da população com base nas âncoras

Fonte: Adaptado de Guyatt *et al.* (2002).

dade de uma população em particular, como o desvio padrão de uma medida, quando aplicada à população em geral num determinado ponto do tempo, e referir-se sempre à mesma medida. A segunda, é escolher o erro padrão de medida, teoricamente independente da amostra. Nenhuma destas soluções, contudo, confronta o segundo problema fundamental dos métodos baseados na distribuição.

Para decidir se a magnitude de um efeito de um tratamento justifica os riscos e os custos associado a esse tratamento, um clínico que saiba, por exemplo, que o efeito é de 0,3 desvio padrão, tenderá a não avançar mais, pois normalmente as unidades de desvio padrão não têm significado intuitivo para os clínicos. No entanto, é possível que possam ganhar experiência com as unidades de desvio padrão da mesma forma que compreendem as pontuações de QdV. Cohen (1988), tendo em conta este problema, sugere que as alterações 0,2 desvios padrão representem pequenas alterações, as de 0,5 desvio padrão representem alterações moderadas e as de 0,8 desvio padrão representem grandes alterações. O problema desta abordagem é a sua arbitrariedade: será que os valores de 0,2, 0,5, e 0,8 desvios padrão representam sempre efeitos pequenos, moderados e grandes?

Na resposta a este problema, investigações recentes tentaram obter provas empíricas sobre a relação entre os resultados dos métodos baseados na distribuição e os baseados numa âncora. Estes estudos abordam a questão «qual é a interpretação apropriada de uma determinada magnitude do efeito, em unidades baseadas na distribuição, como julgados pelos resultados de estudos baseados numa âncora?».

Kazis, Anderson e Meenan (1989) deram exemplos de como os tamanhos do efeito podem ser usados como marcos na interpretação da alteração, examinando os valores gerados por diferentes tratamentos. Samsa *et al.* (1999) apresentaram dados que sugerem que o tamanho de efeito de Cohen (1988) pode ser aplicável na generalidade. Em situações em que uma única escala é utilizada em diversos estudos, pode observar-se a magnitude absoluta da alteração em vez do número de unidades de desvio padrão, o que, em princípio, daria a mesma informação, embora livre do problema das variações dos desvio padrão, no momento inicial, dependentes da heterogeneidade da população.

No entanto, nem todos os investigadores aceitam que a DMS corresponda a uma unidade de 0,2 desvio padrão nos seus estudos. Osoba *et al.* (1998), por exemplo, na sua investigação com doentes oncológicos, defendem que a DMS está entre 0,2 e 0,5. Os métodos baseados na distribuição também têm sido criticados devido ao facto da variabilidade diferir de estudo para estudo, dependendo da heterogeneidade

da população. Uma estratégia para lidar com este problema é usar o erro padrão de medida (EPM), que pode ser definido como a variabilidade entre a pontuação observada de um indivíduo e a verdadeira pontuação. Calcula-se através da multiplicação do desvio padrão no momento inicial e a raiz quadrada de 1 menos a fiabilidade da medida de QdV.

Teoricamente, é importante realçar que o erro padrão de medida é independente da amostra, enquanto que as suas componentes estatísticas — o desvio padrão e a estimativa de fiabilidade — o não são. De facto, à medida que a variabilidade entre-pessoas aumenta numa população em estudo, o desvio padrão irá aumentar (tendendo a aumentar o erro padrão), mas a fiabilidade irá também aumentar (tendendo a baixar o erro padrão). Assim, o EPM reflecte largamente a variabilidade intra-pessoa ao longo do tempo. Wyrwich *et al.* (1999) apresentaram dados comparando o erro padrão de medida com a DMS na utilização do *Chronic Respiratory Disease Questionnaire* e no *Chronic Heart Failure Questionnaire*. Estes autores obtiveram uma correspondência entre a abordagem baseada numa âncora e o critério de 1 erro padrão. Outros investigadores sugeriram que estimativas mais amplas (erro padrão até 2,77) representam alterações importantes nas medidas psicométricas e fisiológicas (Movsas, 2004). Uma decisão crítica no cálculo do erro padrão consiste em se utilizar o método de coerência interna ou o teste-reteste para o cálculo da fiabilidade. Embora Wyrwich *et al.* (1999) argumentem a favor da coerência interna, outros são a favor do teste-reteste.

Há, assim, uma relação entre medidas estatísticas da variabilidade do doente e estimativas baseadas numa âncora de pequenas, moderadas ou grandes diferenças na QdV. Considerando que os desvios padrão ao longo dos estudos da QdV que usam os mesmos instrumentos são coerentes, observar-se-á uma relação entre o desvio padrão e a DMS. Se esta relação fosse também coerente para os vários instrumentos de medição, esta área de investigação tornar-se-ia muito mais fácil. Embora a relação entre medidas de variabilidade e a DMS sejam diferentes (e haverá com certeza grande variabilidade), as conclusões clínicas dos estudos da QdV podem provar serem robustas à variabilidade vista na maior parte dos estudos de QdV. Se assim for, os métodos baseados na distribuição podem tornar-se extremamente úteis (Movsas, 2004).

### 3.3. Decisões com base no indivíduo ou no grupo

Os observadores fazem frequentemente a distinção entre a significância de uma alteração específica



numa pontuação associada a um indivíduo e uma alteração da mesma magnitude na média de uma pontuação de um grupo de doentes. Uma alteração de, por exemplo, 2 mm Hg na média da pressão arterial de uma população seria trivial num indivíduo, mas poderia traduzir-se numa redução do número de trombozes nessa população.

Há duas razões para esta diferença de interpretação. Por um lado, os indivíduos pertencentes a uma mesma população não apresentam exactamente a mesma mudança no resultado; antes, há uma distribuição, isto é, alguns doentes atingem uma redução da pressão arterial muito maior do que a média, enquanto outros atingem menos ou até podem ter um aumento na pressão sanguínea como resultado do tratamento. Considerar a variabilidade na resposta individual ressalta a deficiência fundamental de sumariar os efeitos de tratamentos como uma diferença de médias. O médico que assume que cada indivíduo apresenta o efeito médio é passível de cometer decisões clínicas erróneas e, em alguns casos, com consequências desastrosas. Mais uma vez, dependendo da distribuição das diferenças individuais, a mesma diferença média pode ter implicações distintas. Por exemplo, considere-se a situação em que existe um limiar abaixo do qual qualquer mudança no estado de saúde não terá qualquer consequência importante para os doentes e que a diferença média na população em estudo está abaixo desse limiar. Se a distribuição da alteração com o tratamento é esguia (desvio padrão pequeno), é possível que nenhum doente atinja um benefício importante com o tratamento; por outro lado, se a distribuição da alteração é achatada (desvio padrão grande), um número substancial de doentes poderá atingir esse benefício.

#### 4. Discussão

Segundo Guyatt *et al.* (2002), os responsáveis por recomendações de tratamento — clínicos no caso de doentes individualmente, especialistas e decisores políticos de saúde para grupos de doentes — têm de pesar os benefícios esperados de um tratamento contra os seus efeitos adversos, efeitos tóxicos e custo. Este processo requer um entendimento razoavelmente preciso dos benefícios e riscos dos tratamentos.

No entanto, adquirir este entendimento representa um problema significativo, mesmo para os resultados clínicos dicotómicos como AVC, enfarte do miocárdio ou morte. Por exemplo, a forma como os ensaios clínicos apresentam os resultados condiciona o médico na forma de tratamento. Ainda mais com-

plexo é quando se considera que o valor que os doentes colocam num determinado benefício, poder variar entre indivíduos e ao longo do tempo. Estas dificuldades ocorrem apesar da facilidade com que cada pessoa conceptualiza um acontecimento como um AVC, um enfarte do miocárdio ou a morte. O problema torna-se consideravelmente mais desafiador quando os decisores lidam com os resultados de QdV de uma escala ordinal ou contínua, que lhes é desconhecida.

Mas a complexidade deste assunto é também dependente da perspectiva utilizada, seja ela a perspectiva do doente, do médico ou da sociedade, pois a QdV é um conceito multidimensional que inclui os domínios físico, psicológico, social e emocional (Frost *et al.*, 2002). As medidas de QdV captam o impacto no dia-a-dia da doença e tratamento no funcionamento e na percepção de bem-estar. Uma alteração clínica importante tem sido descrita como uma alteração nos valores da QdV que os doentes percebem como benéfica ou maléfica e importante, e que os pode levar a procurar cuidados médicos, a pedir mudanças no seu tratamento ou a avaliar o êxito de um tratamento. Um declínio de cinco pontos na QdV pode representar um impacto suficiente num indivíduo que o leve a procurar tratamento médico ou terapias alternativas com o objectivo de melhorar a sua QdV (Frost *et al.*, 2002). Assim, pode-se, por exemplo, perguntar ao doente se a sua QdV melhorou, ficou na mesma ou piorou desde a última consulta. Se piorou, também se pode perguntar se a alteração foi suficientemente grande para se proceder a uma alteração no tratamento. Além disso, pode-se ainda perguntar que domínios da QdV é que são mais importantes, e estes seriam assim alvo de uma avaliação mais profunda. Os domínios que não fossem classificados como importantes poderiam ser ignorados no que diz respeito à estimativa da alteração clinicamente significativa.

No entanto, há situações em que um doente poderá não ser capaz de avaliar a QdV ou de fazer uma classificação precisa da alteração clinicamente significativa, como quando é distraído pela fadiga, pela dor grave ou pela náusea. A deficiência cognitiva pode também afectar a capacidade dos doentes em dar informação precisa acerca da maneira como se estão a sentir. Nestes casos, um observador externo pode ser capaz de avaliar a QdV se actuar como o «cuidador» do doente. No entanto, é também necessário ser cauteloso na interpretação das classificações da QdV que os cuidadores fazem. Muitas vezes as classificações doente-cuidador não são coincidentes já que os últimos tendem a subestimar a QdV dos doentes. No geral, a maior concordância verifica-se mais no domínio físico do que no psicossocial.

Outro conceito a ter em conta é o dos padrões inter-

nos quando se captam as percepções dos cuidadores na avaliação da importância clínica das alterações na QdV. Estes padrões definem o que é uma alteração desejável ou não desejável e são usados pelos indivíduos para determinar o tipo e a magnitude da alteração considerada por eles como clinicamente significativa. No entanto, também estes padrões diferem entre os indivíduos e uma condição que faz com que não valha a pena viver, para uma determinada pessoa, pode ser vista, por outra, como um desafio e um obstáculo a transpor. Por outro lado, os cuidadores usam os seus próprios padrões internos quando avaliam a QdV, que, naturalmente, podem, também eles, ser diferentes dos dos doentes. Estas diferenças afectam, subsequentemente, as avaliações da QdV. O defeito cognitivo pode tornar impossível para os doentes preencherem os instrumentos tradicionais de QdV que envolvem escalas de resposta com múltiplas categorias. Não obstante, há técnicas para ajudar os doentes na avaliação da sua própria QdV, em vez de nos basearmos na opinião dos cuidadores. Ajudas visuais como histogramas com barras de tamanho progressivamente maior facilitam a compreensão do doente em relação às escolhas de resposta. Além disso, repartir a tarefa cognitiva em partes mais pequenas pode também ajudar. Por exemplo, doentes com alguma incapacidade cognitiva, com uma pontuação entre 18 e 28 dada pelo *Mini-Mental State Examination* foram capazes de dar respostas válidas numa escala de resposta com seis pontos, usando uma série de perguntas dicotómicas. Foi perguntado inicialmente aos inquiridos se estavam satisfeitos ou insatisfeitos com um determinado aspecto das suas vidas. Depois de isso estabelecido, foi-lhes pedido para classificarem o seu grau de satisfação, tendo três hipóteses de resposta (leve, moderada, ou muito satisfeitos/insatisfeitos). Também foi mostrado um histograma com uma lista de opções para facilitar as suas decisões (Frost *et al.*, 2002).

No entanto, há também quem defenda que a perspectiva do doente deve ser usada principalmente na definição da QdV e que talvez seja menos desejável utilizá-la para determinar a significância clínica. De entre os argumentos invocados surge que esta perspectiva é essencialmente subjectiva e que os dados de QdV são mais significativos quando interpretados em termos de «dados objectivos», como o juízo de um médico acerca da melhoria da condição do doente num parâmetro clínico.

Do ponto de vista do médico, a significância clínica da pontuação da QdV pode ser definida como o efeito mais pequeno que o leva a recomendar um tratamento ou uma terapia ao doente. Quando avaliam a importância de uma alteração na QdV, os médicos tendem a focar-se nos aspectos mais obser-

váveis, como a funcionalidade física, e em indicadores objectivos como, por exemplo, não ser capaz de voltar ao trabalho. Em parte, isto ocorre porque estes aspectos são mais congruentes com os objectivos tradicionais da medicina. Em contraste, os médicos tendem a ter mais incertezas acerca da importância clínica das alterações nas pontuações da QdV. Os médicos experientes hesitam pouco em agir de acordo com as medidas clínicas, com pontuações contínuas, sobre as quais julgam o estado dos seus doentes. Concentração de hemoglobina, contagem de plaquetas, nível de creatinina e capacidade de exercício na passadeira constituem alguns exemplos.

Há, contudo, que questionar de que forma se interpretam estes resultados ou, por exemplo, como é que um médico decide que uma alteração no volume expiratório forçado num segundo (FEV1) de 15% se aproxima de um resultado clinicamente significativo. Normalmente valoriza o grau da diferença necessário para manter o doente fora do hospital, para aumentar a probabilidade de o doente voltar ao trabalho ou tornar o doente menos dispneico enquanto faz as suas actividades de vida diárias básicas. Além disso, é também capaz de observar o doente directamente e perguntar-lhe as suas próprias impressões. A sua experiência com a relação entre o instrumento alvo e as medidas independentes, processadas holisticamente, informam o seu juízo.

Pelo facto de os médicos raramente usarem medidas de QdV na prática clínica, as abordagens intuitivas que os investigadores usaram para estabelecer a interpretabilidade de medidas clínicas são uma opção por desenvolver. Segundo Frost *et al.* (2002), os médicos podem lidar informalmente com questões da QdV na prática clínica, mas desconhecem, muitas vezes, que perguntas são feitas, com que periodicidade são feitas e como é que as respostas são agrupadas. Algumas das abordagens iniciais na medição de QdV dependiam de classificações unidimensionais realizadas pelo médico, como a escala de desempenho de Karnofsky, e da sua familiaridade com as medidas tradicionais biológicas ou físicas, estas são muitas vezes utilizadas informalmente para obter a percepção da QdV do doente.

As medidas clínicas como a pressão arterial e o hemograma estão, de facto, associadas a importantes diferenças nas pontuações de QdV e podem ser usadas para detectar diferenças clinicamente importantes. No entanto, estas medidas podem também levar a conclusões erróneas, quando usadas como único indicador para estabelecer diferenças clinicamente importantes nas pontuações da QdV. Assim, pode não haver uma alteração na QdV de um indivíduo mesmo que haja uma alteração biológica. Inversamente, a QdV de um indivíduo pode também mudar

sem que haja uma alteração biológica. Estas situações ilustram como as medidas clínicas tradicionais podem falhar em captar o impacto idiossincrático da doença e tratamento num doente individual.

Tem também sido sugerido que, devido ao facto de serem os cidadãos a pagar os cuidados médicos, deviam ser os valores da sociedade a determinar as prioridades dos cuidados de saúde. A sociedade tem um interesse inerente em manter uma população saudável e produtiva e uma melhor QdV para o número máximo de pessoas possível. Subgrupos da população também têm interesses na QdV de determinadas subpopulações. Por exemplo, grupos de apoio de cancro trabalham para os interesses de doentes oncológicos, ajudando a compreender melhor as dificuldades com a QdV vivida por estes doentes. Do ponto de vista da sociedade, a QdV é geralmente vista em termos de trocas entre o peso dos recursos utilizados e os benefícios que daí possam derivar, tendo em conta o número de doentes que irá tirar partido desse benefício. A definição de diferenças clínicas importantes do ponto de vista da sociedade é largamente baseada nos valores do grupo estudado, sendo a sua importância definida pelos resultados que são considerados como merecedores dos recursos da sociedade. Por exemplo, pode-se decidir que um determinado tratamento vale o esforço de gastar os recursos da sociedade, se existir uma probabilidade alta de fazer com que mais pessoas regressem ao trabalho a tempo inteiro (Frost *et al.*, 2002).

É possível que, ao longo do tempo, os médicos se tornem mais familiarizados com os parâmetros da significância clínica da QdV (como a DMS, o NNT ou o EPM), da mesma forma que estão já familiarizados com os valores estatísticos de  $p$  (Movsas, 2004). Até porque é importante ter em conta que as diferenças na interpretação da significância clínica da QdV dependem da perspectiva particular do observador, seja ele doente, médico ou decisor (Osoba 1998). Por outro lado, a DMS é um objectivo final para o qual nenhum valor limiar é universalmente aceite. Além disso, se é observada uma diferença num objectivo ou resultado final de QdV que não preenche os critérios da DMS, não se devia simplesmente ignorar esse achado pois, por definição, a QdV vai para além dos objectivos ou resultados finais clínicos padronizados. Assim, poderá haver achados de QdV importantes que nunca serão directamente correlacionados com um parâmetro clínico. Em vez disso, é necessário avaliar cuidadosamente se o resultado de QdV é apenas um fenómeno estatístico *versus* um fenómeno real, devendo a observação ser usada para gerar uma hipótese a ser testada num futuro estudo (Movsas, 2004).

## 5. Conclusão

Foi nosso objectivo, não o estudo exaustivo do tema, mas antes, verificar, através de uma revisão de trabalhos anteriores, o conceito de significância clínica e a importância da sua determinação ou medição na prática clínica diária em oncologia, de modo a medir com rigor as alterações globais da QdV do doente oncológico.

Cada vez mais se compreende que embora a sobrevivência e a cura sejam factores críticos para os doentes oncológicos, a QdV global é também fundamental (Bottomley, 2002). Assim, é imperativo que a sua avaliação faça parte da prática clínica no dia-a-dia. O doente é quem melhor pode explicar o que sente e que alterações clinicamente significativas ocorreram para que, em conjunto com o médico, se proceda ou não a uma alteração dos cuidados médicos.

É necessária, no entanto, mais pesquisa nesta área para capacitar profissionais de saúde de directrizes para interpretar a significância clínica. Embora se defenda que a QdV deva ser definida com base na perspectiva do doente, a realidade é que continua a ser definida pelo médico em termos de acontecimentos observáveis como a capacidade de voltar ao trabalho e a ausência de sintomas. Assim, é necessário um papel mais proeminente do relato de QdV do doente, assim como a interpretação do que constitui uma alteração clinicamente significativa.

Espera-se que este artigo possa promover e incentivar a investigação deste tema, por parte dos autores portugueses, bem como a aplicação do conceito de significância clínica na tomada de decisões médicas na prática clínica do dia-a-dia.

## □ Referências bibliográficas

- BOTTOMLEY, A. — The cancer patient and quality of life. *The Oncologist*. 7 : 2 (2002) 120-125.
- CELLA, D. *et al.* — Advances in quality of life measurements in oncology patients. *Seminars in Oncology*. 29 : 3 (Suppl 8) (2002) 60-68.
- COHEN, J. — Statistical power analysis for the behavioral sciences. 2<sup>nd</sup> ed. Hillsdale, NJ : Lawrence Erlbaum Associates, 1988.
- FROST, M. *et al.* — Patient, clinician, and population perspectives on determining the clinical significance of quality-of-life scores. *Mayo Clinic Proceedings*. 77 : 5 (2002) 488-494.



GUYATT, G. *et al.* — Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings*. 77 : 4 (2002) 371-383.

JAESCHKE, R.; SINGER, J.; GUYATT, G. H. — Measurement of health status : ascertaining the minimal clinically important difference. *Controlled Clinical Trials*. 10 : 4 (1989) 407-415.

KAZIS, L. E.; ANDERSON, J. J.; MEENAN, R. F. — Effect sizes for interpreting changes in health status. *Medical Care*. 27 : 3 Suppl. (1989) S178-189.

KING, M. — The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Quality of Life Research*. 5 : 6 (1996) 555-567.

MOVSAS, B. — Quality of life in oncology trials : a clinical guide. *Seminars in Radiation Oncology*. 13 : 3 (2004) 235-247.

OSOBA, D. *et al.* — Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology*. 16 : 1 (1998) 139-144.

OSOBA, D. — Interpreting the meaningfulness of changes in health-related quality-of-life scores : lessons from studies in adults. *International Journal of Cancer*. Supplement. 12 (1999) 132-137.

RODRIGUES, G. *et al.* — The relationship of changes in EORTC QLQ-C30 scores to ratings on the Subjective Significance Questionnaire in men with localized prostate cancer. *Quality of Life Research*. 13 : 7 (2004) 1235-1246.

SAMSA, G. *et al.* — Determining clinically important differences in health status measures : a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics*. 15 : 2 (1999) 141-155.

SLOAN, J. A.; DUECK, A. — Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. *Journal of Biopharmaceutical Statistics*. 14 : 1 (2004) 73-96.

SLOAN, J. A. *et al.* — Assessing clinical significance in measuring oncology patient quality of life : introduction to the symposium, content overview, and definition of terms. *Mayo Clinic Proceedings*. 77 : 4 (2002) 367-370.

SPIPKER, B. — Quality of life and pharmacoeconomics in clinical trials. 2<sup>nd</sup> ed. New York : Lippincott-Raven, 1996.

WARE, J. E.; KELLER, S. D. — Interpreting general health measures. In SPIPKER, B. ed. lit. — Quality of life and pharmacoeconomics in clinical trials. 2<sup>nd</sup> ed. Philadelphia : Lippincott-Raven, 1996. 445-460.

WYRWICK, K. *et al.* — Linking clinical relevance and statistical significance in evaluating intraindividual changes in health-related quality of life. *Medical Care*. 37 : 5 (1999) 469-478.

## □ Abstract

### CLINICAL SIGNIFICANCE: A REVIEW

Quality of life (QoL) is becoming more and more a relevant indicator of how the patient and the clinician handle oncological disease. So, it is necessary to criteriously understand and evaluate the changes in QoL throughout the treatment. As long as the physician has a rigorous and stable measurement indicator, s/he can assess if any change that has occurred in the QoL is significant or not; if it justifies a change in the therapeutics plan or in the clinical decision making. The interpretation of the results of QoL studies has been scrutinized by investigators because statistical significant results (not a product of chance) are not necessarily clinically significant. This means one needs a measurement to identify the minimal clinical significance, that patient and clinician have to perceive, in order to proceed to changes on the therapeutic plan and/or daily activities. The main goal of this article is a revision of the literature on the clinical significance concept in oncology. We hope that, in Portugal, it may contribute to the promotion of a better understanding and research on the theme, as well as to its use in daily practice, in order to promote the best QoL to the oncology patient.

Keywords: oncology; quality of life; clinical significance.



Preço de capa, 15 €